图书情报工作
第63卷 第1期 2019年1月

从信息检索系统评估到知识服务平台评估

■ 李月琳 韩宏亮
南开大学商学院信息资源管理系 天津 300071

摘要：［目的/意义］随着知识服务平台的发展，如何通过评估改善知识服务平台的功能和设计以更好地服务于用户成为重要的问题。本文旨在探讨知识服务平台的评估要素和评估标准，为知识服务平台的评估实践提供理论指导。［方法/过程］从信息检索系统评估出发，回顾和梳理信息检索系统评估、数字图书馆评估的相关研究和理论模型、数字图书馆的评估要素和标准。［结果/结论］基于总结和归纳以往信息服务类系统的评估要素，结合知识服务平台的特征，阐述知识服务平台评估的要素和评估标准。

关键词：知识服务 知识服务平台评估 理论框架
分类号：G235.76
DOI：10.13266/j.issn.0252-3116.2019.01.008

1 引言

随着数字图书馆的发展以及知识服务需求的增长，越来越多的数字图书馆超越了以往信息服务的范畴，迈向知识服务领域，并从服务内容和服务手段方面实现了对传统信息检索系统和数字图书馆的延伸和拓展，形成了新型的知识服务平台。国内知名的数字图书馆，包括CNKI、万方和维普，都在近年来增加了新的服务功能，明确了新的知识服务方向。如新版的CNKI提供"行业知识服务和知识管理平台""研究学习平台""专题知识库"等板块，同时提供不同角度的文献计量分析，从而不仅为用户提供文献的信息服务，同时基于对文献的分析和挖掘，为用户提供不同层面的知识服务。此外，为更广泛地为用户提供服务，包括企、事业单位等不同的机构和组织提供知识服务也成为知识服务平台的重要任务。国外方面，Wiley和Elsevier等传统的数字图书馆均在近年来调整定位，开展旨在为用户提供学术研究、专业领域支持等多方面的知识服务；Elsevier还拓展其医疗知识服务，同时为用户提供文献分析、趋势预测、作者服务等多项增值服务[1]。国内外的发展趋势均表明知识服务平台将成为新型的信息和知识服务系统。然而，相对于蓬勃发展的知识服务平台的开发和建设，如何评估其功能和绩效？尤其是结合知识服务的目的和特点开展知识服务平台的评估？学术界依然鲜有探讨。因而，本文将探讨这一问题，试图为后续的知识服务平台评估理论研究和实践发展提供借鉴。

本文将知识服务平台定位为信息服务类系统，从传统信息检索系统评估入手，回顾数字图书馆的评估理论模型，并从中析出知识服务平台的评估要素，结合知识服务的特点，阐述知识服务平台评估的各要素和评估标准。

2 早期的信息检索系统评估

信息检索系统评估是发现系统不足，完善系统功能设计，满足用户信息需求，提升用户体验，进而提升信息服务质量的重要途径。信息检索系统是一个不断发展的概念，广义上来看，信息检索系统包括：传统的文献数据库，如早期的Dialog系统；数字图书馆，如我国的CNKI、美国的ACM数字图书馆等；互联网环境下诞生的网络搜索引擎，如Google、Baidu系统等；顺应知识管理和服务需求逐步发展起来的知识服务平台。对这些不同类型的信息检索系统的评估，是信息服务类系统开发和建设的需要，也是系统开发、建设和不断完善过程中必不可少的环节。早期最著名的信息检索系统评估当属Cranfield检测（Cranfield tests）。Cranfield检测是由C. W. Cleverdon主持的、长达10余年的信息检索系统评估项目。前期研究开始于1953年[2]，是Cranfield 1的预实验，即采用单元词标引航空文献。此

作者简介：李月琳（ORCID：0000-0002-1496-6741），副院长，系主任，教授，博士生导师，E-mail：yuelinli70@163.com；韩宏亮（ORCID：0000-0001-5080-9244），博士研究生。
收稿日期：2018-10-16 修回日期：2018-12-03 本文起止页码：52-59 本文责任编辑：王传清

后开展的正式研究分两个阶段完成，包括 1957 - 1961 年的 Cranfield 1 及 1963 - 1966 年的 Cranfield 2。Cranfield 1 的目的是研究和比较 4 种标引的效率，包括通用十进制分类法（The Universal Decimal Classification，UDC）、字顺主题词表（Alphabetical subject catalogue）、分面分类体系（A faceted classification scheme）及组配标引的单元词系统（The uniterm system of coordinate indexing）。Cranfield 1 的实验结果表明了信息检索（information retrieval，IR）系统评估的必要性和可行性，初步确立了系统评估的方法和模式。同时，揭示了查全率与查准率之间的关系，其研究结果为 Cranfield 2 的开展奠定了基础。Cranfield 2 扩大了样本量，尽可能采用接近真实的检索问题，同时引入了相关性判断，以寻求更可靠和有效的研究结果。总之，历时 10 余年的 Cranfield 检测留下了丰富的学术资产，深刻地影响了 IR 系统评估，在情报学发展的历史上留下了浓墨重彩的一笔。K. S. Jones[2] 总结了 Cranfield 检测的贡献，包括：检测证明了不同的标引语言（包括自然语言）的绩效几乎没有区别，对人们理解和认识信息检索系统产生了积极和正面的影响。更为重要的是，Cranfield 实验从方法论上明确了信息检索实验的标准，确立了测试集的概念及实验的流程、评价指标等，为此后系统导向的 IR 系统评估的开展奠定了理论和实践基础。其确立的方法体系至今仍然在 Text Retrieval Conference（TREC）实验中被广泛运用。

TREC 会议始于 1992 年，是美国从政府层面试图推动产、学、研结合而推出的重要学术会议，目前已发展成全球性的国际学术会议，并吸引了不同的国家和地区开展相应的学术研究活动，试图通过利用和传播在信息检索领域的学术研究成果促进社会对信息和信息技术的运用。第一届 TREC 会议于 1992 年 11 月在马里兰州的 Gaithersburg 召开。最初的 TREC 会议关注的内容相对比较狭窄，聚焦于文本信息的检索。近年来，随着网络信息资源的不断丰富及信息检索系统概念的扩展，TREC 会议不断地调整会议主题，跟踪前沿领域，拓展研究范围，会议的内容和形式日益丰富。除传统的文档资料外，TREC 会议着眼于多种信息载体类型的检索研究，如语音、视频。此外 TREC 还关注非英语文档检索和跨语言检索。TREC 会议通过评估的方式不断推动信息检索技术水平的提升和改善。TREC 会议所建立的评估范式和各种测试集极大地影响了信息检索系统技术及评估的发展。如今，TREC 会议已经成为信息检索领域最重要的会议之一。

早期信息检索系统评估研究中，不同于系统导向特征明显的 Cranfield 检测，MEDLARS 系统的绩效评估则体现了用户导向的 IR 系统评估的特点。MEDLARS 评估的目的包括：① 研究该系统用户的搜索要求；② 探究 MEDLARS 服务如何高效和有效地满足这些要求；③ 识别影响系统绩效的不利因素；④ 揭示如何更有效和经济地满足 MEDLARS 用户搜索要求的方法，尤其是为如何更有效地利用新一代的设备和程序提出建议[3]。评估中使用的检测请求（test requests）来源于用户的真实需求，同时兼顾用户搜索主题的代表性，即搜索主题涵盖 MEDLARS 系统所收录的不同领域的内容。用户方面，选择了 20 个医疗组织形成了用户群，并邀请他们参与该项评估。然而，评估指标方面依然着重考察查全率和查准率，并详细地分析了查全率和查准率失败的原因，同时指出了用户与系统的交互障碍是导致失败的重要原因之一。MEDLARS 系统评估是早期的评估研究中不多见的将用户纳入考虑的评估项目，尤其是对用户与系统交互的分析，为此后用户导向的信息检索系统评估实践和发展奠定了基础。

3 数字图书馆评估理论模型

随着传统的信息检索系统向数字图书馆方向的发展，数字图书馆的评估成为 IR 评估领域的焦点。在数字图书馆评估领域，系统导向的评估依然占据主导地位[4]。然而，用户导向的数字图书馆评估因其能更有效地支持数字图书馆的改善和服务能力的提升而受到重视。

T. Saracevic[5] 曾指出，数字图书馆的评估滞后于蓬勃发展的数字图书馆的建设和实践。为此，基于对传统信息检索系统研究的回顾，他讨论了数字图书馆的概念、评估的内容、评估的标准、评估的方法及评估数字图书馆的原因，提出了一个帮助我们思考如何评估数字图书馆的框架[5]，涵盖评估的要素和关键问题，具体内容包括：① 评估的概念：评估什么？什么是数字图书馆？包含什么内容？评估包括哪些要素？②评估的环境：选择评估的目标、框架、观点或者选择评估的层级。评估包括哪些层次？对评估的某个层次来说哪些是关键的问题？最后，不同层次的评估目的是什么？③评估的标准：标准反映了与评估目的相关的绩效表现。评估应聚焦于何种绩效参数？评估何种维度和特点？④测度：测度指标反映选择的标准以记录系统绩效。评估过程中针对选择的标准，应采用哪些具

体的测度指标？⑤评估的方法：使用哪些测量手段和工具？如何取样？数据收集采用什么程序？数据分析采用什么方法？

T. Saracevic 所提出的框架为数字图书馆评估的理论和实践研究提供了指导。

交互是用户获取系统资源的重要手段。N. Fuhr 等[6]提出了一个以交互为核心的数字图书馆评估三维模型，三维指用户、内容和系统三大交互要素及它们之间的关系。他们指出，数字图书馆的评估应该关注这三个要素，并着重评估三者之间的关系：用户与内容之间的评估重点是有用性；用户与系统之间的评估应注重可用性；系统与内容之间应着重评估绩效。G. Tsakonas 等[7]进一步阐述了三维框架，分析了用户、系统和内容蕴含的关键要素，并提出了有用性、可用性和绩效评估的标准和子标准。针对有用性，目标属性和资源属性是评估的标准；可用性以有效性、效率和用户满意度为评估的标准；查准率、查全率、相关性和反应时间是评估系统的主要标准。这些标准的提出丰富和完善了 N. Fuhr 等提出的数字图书馆评估的三维模型，也为思考如何评估信息服务类系统提供了借鉴。

基于 T. Saracevic 的评估框架及分层交互模型[8]，Y. Zhang[9]从数字图书馆不同的利益相关者的视角出发，开展数字图书馆的评估研究，提出了一个整体性的数字图书馆评估模型。该模型强调：不同评估标准的重要程度对不同的利益相关者而言存在差异，即不同的利益相关者针对数字图书馆的不同方面，如环境、技术、用户、界面、内容和服务进行评估时，往往选择不同的评估标准。如用户在评估界面时，仅考虑需付出的努力和交互支持，然而，在评估技术时，绩效和效率则是他们选择的评估标准。Y. Zhang 的研究启示我们数字图书馆的整体性评估可从数字图书馆的不同要素出发加以考虑，同时要考虑不同利益相关者的需求。

李月琳等[10]的研究则在以往研究的基础上，提出有效的数字图书馆交互功能的设计是数字图书馆成功的保障的观点。因而，基于交互这一核心要素评估数字图书馆是必要的。为此，他们将任务融入用户的交互模型中，构建了数字图书馆的三维交互模型。三维即指任务、数字图书馆的支持技术及信息资源，而用户是交互的核心。他们进一步探讨不同的交互维度如何影响用户从不同视角对数字图书馆的评估。通过实验研究方法，基于用户与数字图书馆交互，他们提出了数字图书馆交互功能评估模型、基于任务的数字图书馆评估模型和整体性数字图书馆评估模型。研究发现，针对用户评估数字图书馆视角的不同，或评估目的的不同，评估的标准和要素也呈现出差异性，且不同的标准发挥的作用不同。这些评估模型和评估标准的提出与验证，及对不同的标准在评估中发挥作用的探讨，为进一步构建知识服务平台评估框架提供了方法。

尽管数字图书馆和知识服务平台在服务目的、提供的最终产品、针对的用户需求、服务形式等方面各有侧重，然而，知识服务平台体现了数字图书馆的诸多特征，包括系统界面、人机交互功能设计、信息与知识内容的获取、组织与提供等，是数字图书馆的延伸和发展。因而，数字图书馆评估的理论、思路和方法可以为知识服务平台评估提供参考和借鉴。

4 知识服务平台的评估

如同 T. Saracevic[5]所指出的，数字图书馆评估的发展滞后于其实践的发展。知识服务平台近年来蓬勃发展，然而，如前所述，知识服务平台评估的研究和发展却相对滞后。本文试图沿着 T. Saracevic[5]探讨数字图书馆评估的路径，分析和构建知识服务平台的评估模型，即从“什么是知识服务平台?”入手，探讨“为什么要评估知识服务平台?”“评估什么?”“评估的标准是什么?”及“如何在评估中应用这些标准?”

4.1 什么是知识服务平台?

无疑，知识服务平台是提供知识服务的信息服务类系统。不同于传统的信息检索系统、搜索引擎和数字图书馆仅以提供信息为其服务的主要内容，知识服务平台增加了知识服务这一核心内容。传统的信息服务以信息搜集、整理、传递等信息工作为主要内容，知识产品的提供通常不是信息服务机构的工作范畴；知识服务则以提供问题的知识解决方案为核心，帮助用户解决他们的问题。正如张晓林[11]指出的，信息服务着重于“我是否提供了您需要的信息?”，而知识服务则关注“是否通过我的服务解决了您的问题?”，由此可见信息服务与知识服务的区别。所以，知识服务平台评估的核心问题应该有别于数字图书馆，更聚焦于知识服务平台是否帮助用户实现了“问题解决”的初衷。

如前所述，知识服务平台一方面是数字图书馆的拓展和延伸，即在数字图书馆的信息服务基础上，增加了知识服务的功能，实现对原有数字图书馆功能的提升和改造；另一方面，一些知识服务平台从一开始便确定了知识服务的核心功能，以知识服务为主、信息服务为辅的建设方针，着力打造新型的信息服务类系统。

目前知识服务平台可以分为三种类型：其一是企业依据自身状态及需求，整合行业资源构建的面向某一行业的知识服务平台，如LOTUS、微软和清华同方开发的面向企业的知识管理平台，面向科技型中小企业的企业知识服务平台等。其二是以信息服务类系统为基础，在此基础上拓展新的知识服务功能、基于数据库提供知识检索服务的数据库类型平台。这类知识服务平台的衍生以数字图书馆及各大出版商构建的索引数据库为代表。如国内的中国知网、万方、维普，国外的Web of Science、Emerald、Elsevier等。其三是面向社会公众依托互联网进行知识共享的百科类及问答式知识服务平台，如Wikipedia、Quora、知乎、得到等。

4.2 为什么要评估知识服务平台？

如同建设数字图书馆，知识服务平台也需投入大量的资金加以建设和维护，包括软硬件的建设和维护，信息内容的购买和获取，人员的招募和培训等。如此大量的资金投入，其社会效益和经济效益如何？是否达到了预期的效果？为回答这些问题，评估知识服务平台是十分必要的。基于知识服务平台性质的不同，对某些具有公益性质的知识服务平台而言，社会效益也可能是更为重要且需要着重评估的方面。

在技术层面上，知识服务平台提供的技术支持是否有效地提供了知识服务也是需要评估的问题。随着信息技术、大数据技术和人工智能技术的不断发展和进步，在信息搜索、知识创造和创新、信息与知识呈现方式上必将越来越多地利用这些新的技术、手段和方法。然而，这些新技术多大程度上可以帮助我们实现知识服务的目的？其在知识服务方面的能力如何却是需要评估的。信息检索系统及数字图书馆评估发展的历史启示我们，只有通过评估，才能发现系统和技术存在的不足和问题，才能为进一步完善知识服务平台的技术支持提供指导。

此外，知识服务平台的主要服务对象是用户，包括个体和机构用户。知识服务平台提供的服务怎样？是否满足了用户的知识需求？是否能帮助用户有效地解决问题？这些问题的回答，一方面可以帮助了解知识服务平台是否达到预期的目的；另一方面，更为重要的是可以从用户的视角帮助知识服务平台丰富内容建设、改善系统设计、提高服务质量。对信息检索系统评估的理论和实践的长期探究，催生了用户导向的信息系统评估，确立了以用户为核心的评估理念，有效地帮助信息检索系统改善服务和与用户的交互，为提升信息检索系统的绩效作出了贡献。对于知识服务平台的评估而言，评估技术能力和系统绩效的同时，关注其对用户需求的支持，采用用户导向的评估途径无疑也是十分重要的。

4.3 评估什么？

早期信息检索系统评估和数字图书馆评估无疑为知识服务平台的评估奠定了基础。知识服务平台作为信息服务类系统与前两者具有很多相似性，如包括技术、内容、界面、服务、交互、和用户等要素。评估也应从不同的要素切入，实现对知识服务平台的局部或整体性评估。

4.3.1 技术　知识服务平台在信息服务系统的基础上融入了许多重要的新技术，如通过大数据技术改善知识服务平台的运行和管理[12]；通过网格技术改进信息资源发现，建立知识网络[13－14]；基于本体技术构建面向专业服务的知识服务平台[15]；利用多语言信息发现技术和异构数据库整合技术集成信息资源；使用推送技术、数据挖掘技术、导航库技术和智能代理技术实现用户需求的分析和个性化、定制化服务等。将如此众多的技术整合在一起，如何实现最佳的系统绩效并使投入产出最大化，是研究和实践界面临的基本问题，也是评估的意义所在。

4.3.2 内容　内容是满足用户需求的核心。知识服务要提供为用户解决问题的方案，知识服务平台需要将网络上的各种信息资源集成整合，形成知识产品。不同的知识服务平台因其所面向的用户群体的差异，在提供的内容上也各有侧重。由数字图书馆延伸而来的知识服务平台在原有功能的基础上，建设面向特定学科专业的知识仓库，构建基于自身和网络资源的知识网络，根据用户自身特点提供个性化、定制化内容[16]；公共知识服务平台面向的用户是普通大众，其提供的内容十分广泛，如质量监督、社会保障、医疗卫生、住房、教育、食品、交通、法律和金融、就业与创业等方面的知识内容，囊括了衣食住行及工作、生活的方方面面[17]；面向特定行业的知识服务平台为专业用户解决问题提供更具针对性、专业性的内容。如在医学行业中，知识服务已经突破传统医学知识库的方式，形成了以Clinical Evidence、DynaMed、Essential Evidence Plus、MD Consult、UpToDate、ClinicalKey等为代表的医学知识服务平台，其所面向的用户涵盖学生、教师及医患双方，在权威医学数据的基础上提供教育、患者交流、问题求解服务，同时支持疾病诊断、治疗、疾病控制和预防等多种临床活动[18－19]。然而，如此丰富的内容能否满足用户的需求？是否能帮助用户解决他们的问

题？能否为用户带来新知或带来启发，帮助他们找到问题解决的方案？为了回答这些问题，开展知识服务平台的内容评估无疑是必要的。

4.3.3 界面 用户界面一直是信息系统设计和系统评估所关注的重点。N. J. Belkin 等通过对欧洲空间研究所的4个信息系统界面项目进行研究，提出信息系统界面设计的导向要以用户目标、任务和特征为基准[20]。李月琳等[10]对数字图书馆界面的评价标准进行了梳理和研究，提出了影响用户与界面交互效果的关联性和预测性交互子维度，并将界面视为用户与技术交互的主要场所。知识服务平台的系统界面也扮演着同样的角色，是沟通用户与系统内容的桥梁。因而，界面的功能设计是用户能否获取有用的知识以解决他们的问题的关键。然而，如何才能为用户提供可用、有效和易用的交互界面？一方面要考虑技术手段；另一方面，用户的体验和感知也应是界面设计考虑的重要因素，而评估是获知这些因素的重要手段。

4.3.4 服务 服务是知识服务平台的核心能力和基本要求。目前在实践应用中的知识服务平台存在不同的类型，其服务模式也各有特点，基于传统数字图书馆衍生而来的知识服务平台在服务模式上继承了已有的规范和体系，包括用户完成检索过程的自主服务模式、专业化服务模式和个性化服务模式，自主服务模式意味着用户独自完成检索过程；专业化服务模式通过参考咨询服务、学科化服务和门站式服务体现知识服务职能；个性化服务模式则是根据用户需求和特点向用户推送相关内容[21-22]。基于互联网而诞生的知识服务平台如"知乎""得到"等则采用基于用户的知识共享方式提供服务，基于海量的数据资源，通过知识推送、订阅、学习、检索等功能为用户提供知识服务[23]。这些服务的模式是否满足了用户的需求？是否为用户提供了更便捷、高质量、人性化的服务则需要通过评估来实现。

4.3.5 用户 用户是知识服务平台面向的对象，在以往的评估实践中，通过对用户行为、特征、任务完成情况、使用及接受意愿的测量实现对系统的评估。李月琳等[10]的研究基于用户的感知，提出了影响数字图书馆的交互功能评估的关联性和预测性子维度。常红等[24]以技术接受模型和任务技术适配模型为基础，构建信息推送用户接受模型，并通过实证研究确定影响图书馆信息推送服务用户接受的因素，提出信息推送服务模式的改进建议。刘明珠和杨建林[25]基于用户的7种信息需求对新浪微博搜索的效果进行评估，通过与百度搜索进行对比，提出微博搜索在检索专业性网站、知识信息及时间成本上的缺陷。此外，部分研究立足用户体验，通过测量用户的使用感受对健康类搜索引擎[26]、微博信息质量[27]、政府网站绩效[28]、界面可用性[29]进行评价。这些研究及以往的基于用户导向的信息检索系统评估研究为如何在知识服务平台情境下开展用户评估提供了有益的启示。

4.3.6 交互 相较于传统的信息系统，交互性在知识服务平台的评估中显得尤为重要。其一是因为科学技术的发展进一步拓展了人与机器的交互方式，语音检索、拍照检索等新的检索方式为用户提供了更多的检索途径，基于AR和VR技术的信息检索、浏览、学习和娱乐已经开始应用于实践。新技术应用于提供知识资源的知识服务平台是一种趋势也是一种必然；其二，由于实践的需求，学界对于人机交互的研究不断增强。评估研究中系统视角向用户视角的转变不断深化，用户成为系统建造、设计、服务的核心，而今用户已经从被动适应转变为主动要求的角色，系统交互性的好坏直接影响用户的使用体验甚至使用意愿。因此，对交互性的评价理应成为知识服务平台评估中的重要方面。

4.4 评估的标准是什么？

通过借鉴信息检索系统评估和数字图书馆评估的标准和维度，结合知识服务平台的任务和目的，笔者构建了针对知识服务平台的评估要素、评估标准和测量指标如表1所示：

表1 知识服务平台评估的要素和标准

评估要素	评估标准	测量指标
技术	有效性	平均查准率、相对查全率、浏览效果
	智能化	智能程度
	效率	响应时间
内容	有用性	帮助问题解决、帮助获取新知、带来启发和灵感
界面	可用性	易用性、易学性、容错能力、易记忆性、效果、效率
服务	质量	满意度、持续使用意愿、推荐意愿
用户	任务完成	完成时间、完成质量
	问题解决	有效性、针对性、启发性
交互	支持程度	用户友好性、需付出的努力
	满意度	

表1中的评估要素来源于对信息服务类系统评估的常用要素[5,9]。评估标准中的有效性和效率是传统信息检索系统评估的常用指标；有用性、可用性是数字图书馆评估的重要指标[9-10]；任务完成程度、交互支持程度、满意度则是基于任务的数字图书馆评估的重要指标[10,30]。问题的解决程度是知识服务平台的目的

所决定的。如前所述,知识服务不同于传统的信息服务,以解决用户的问题为其主旨[11]。因而,评估知识服务平台的绩效加入"问题解决"程度的度量标准和评价指标是必要的。再有,对服务质量的关注来源于长期以来信息服务机构及系统对服务质量的重视,将这一理念移植到知识服务平台评估也是顺理成章的。智能化指标用以衡量人工智能技术在知识服务平台的应用和智能知识服务的实现程度。该指标的纳入是对蓬勃发展的人工智能技术及应用的回应,知识服务平台植入智能化的服务性能是必然的发展,因而,评估也应考虑知识服务平台的智能化程度。

具体测量指标一方面继承了信息检索系统绩效评价指标(包括平均查准率、相对查全率、浏览效果)和可用性评价指标(包括易用性、易学性)等;另一方面,结合知识服务平台的特点及各种功能的实现要求、问题解决程度、任务完成质量、用户满意程度等提出测量指标。如知识服务平台系统强调对问题的解决,因而需着眼于问题解决的有效性、针对性和启发性;对内容有用性的评估也应围绕问题解决的程度,包括帮助问题解决、帮助获取新知、带来启发和灵感的程度。

这些标准和测量指标来源于传统信息检索系统评估、数字图书馆评估,尤其是用户导向的系统评估指标,同时结合了对知识服务平台的目的和宗旨的解读,其能否适用于知识服务平台的评估仍需进一步的检验。

4.5 如何在评估中应用这些标准?

类比数字图书馆评估,知识服务平台评估同样可分为整体性评估和局部评估两种。整体性评估的内容应包括表1中的所有评估要素;局部性评估则关注一个或几个部分并进行更为深入的探究和考察。另一方面,这些标准和测量指标可作为知识服务平台的基础评估指标,具体的评估实践中应依据评估对象的特性给予适当调整,根据需求增删或细化标准或测量指标。具体测量指标的选择应立足于平台所服务的具体人群和范围,着重关注具体的用户任务与需求,以为用户提供解决方案为基本导向,结合平台的宗旨,考虑评估指标的科学性和适用性。同时,要充分利用定性和定量的数据收集和分析方式,注重运用混合数据处理方法,结合定性和定量数据分析解释评估结果。此外,表1的指标体系还可应用于对知识服务平台评估工具的开发,进一步通过实证研究确定评估标准和测量指标的有效性,开发评估量表,确立完整的知识服务平台评估体系。

评估方法方面,实验研究方法是信息系统评估常用的研究方法,从 Cranfield 检测到历时 20 余年的 TREC 会议,实验研究方法都扮演了重要的角色。问卷调查也是可选择的方法之一。为使评估更为有效,问卷调查后的跟踪访谈也是评估中常用的技巧。无论是进行实验研究还是问卷调查,都应注意对真实用户和真实情境的选择。当前系统评估研究的趋势是由实验室实验转向实地实验(field experiment),目的就是为了让用户在最真实的情境和状态下呈现出最本质、自然的认知和行为特点。在具体的操作中,可通过任务来激发用户的信息需求,方式之一是使用用户现实生活中的真实工作或学习任务,这种方式所获得的数据真实可靠,但是也存在着多个用户需求各异,不利于比较和分析的弊端;另一种方式是运用模拟仿真工作任务情境[31],此类任务通常来源于用户的实际需求,由研究人员进行规范化设计并设定具体情境,模拟仿真工作任务情境的有效性在不同研究中得到了验证,是交互信息检索常用的任务设计方法[32-33]。以上两种方式,在研究中或可采用其中一种,或可两者结合,具体依研究的目的而定。此外,随着大数据技术的发展和应用,数据挖掘、情感分析等技术日臻成熟,如何利用用户使用知识服务平台、与平台交互过程中留下的"足迹",即大量的数据和行为轨迹开展评估研究,是值得深入研究的课题。

5 结语

本文基于知识服务平台是信息服务类系统这一基本观点出发,从早期信息检索系统的评估切入,回顾了系统评估的历程及方法的形成,介绍了数字图书馆评估的理论模型、要素和主要标准,并基于 T. Saracevic 的评估路径,结合信息检索系统和数字图书馆评估,论述了知识服务平台评估的要素和标准。

相较于传统的信息检索系统和数字图书馆评估,知识服务平台的评估体现了诸多的不同之处。传统上我们评估"信息"检索系统,而今迈入评估"知识"服务平台,从评估信息检索的效率和绩效到评估知识点、知识单元或模块检索的绩效和效率;从评估信息获取的自动化程度到评估知识获取的智能化程度;从评估信息检索系统中零散的知识服务到评估知识服务平台的集成化的知识服务;从评估普遍的知识提供方式到评估更灵活、个性化的知识提供方式;从评估信息检索系统是否提供了支持用户完成其任务的信息资源到评估知识服务平台是否提供了支持用户解决任务相关问题

和完成任务的知识或方案。这些不同,蕴含着人类在信息处理和获取方面的进一步提升和飞越,如何使这些不同更具现实意义,需要在知识服务平台开发和应用实践中不断开展评估工作,以为其不断完善和提升提供理论和实践支持。从这一角度来看,知识服务平台的评估无疑是至关重要的,我们依然任重道远。

参考文献:

[1] 郝梦. 知识构建视角下的知识服务平台评价研究[D]. 天津:南开大学,2018.

[2] SPARCK J K. Information retrieval experiment [M]. London:Butterworths. 1981.

[3] LANCASTER F W. MEDLARS: report on the evaluation of its operating efficiency [J]. American documentation, 1969,20(2):119-142.

[4] TSAKONAS G, MITRELIS A, PAPACHRISTOPOULOS L, et al. An exploration of the digital library evaluation literature based on an ontological representation [J]. Journal of the Association for Information Science and Technology, 2013,64(9):1914-1926.

[5] SARACEVIC T. Digital library evaluation: toward an evolution of concepts [J]. Library trends, 2000,49(3):350-369.

[6] FUHR N, HANSEN P, MABE M, et al. Digital libraries: a generic classification and evaluation scheme[C]// European conference on research and advanced technology for digital libraries. Berlin: Springer Press, 2001:187-199.

[7] TSAKONAS G, KAPIDAKIS S, PAPACHRISTOPOULOS C. Evaluation of user interaction in digital libraries[OL]. [2018-10-27]. http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.101.9642&rep=rep1&type=pdf.

[8] SARACEVIC T. Modeling interaction in information retrieval (IR): a review and proposal[J]. Proceedings of the American Society for Information Science and Technology, 1996,33(1):3-9.

[9] ZHANG Y. Developing a holistic for digital library evaluation[J]. Journal of the American Society for Information Science and Technology, 2010, 61(1):88-110.

[10] 李月琳, 梁娜, 齐雪. 从交互维度到交互功能:构建数字图书馆交互评估理论模型[J]. 中国图书馆学报, 2016,42(1):66-82.

[11] 张晓林. 走向知识服务:寻找新世纪图书情报工作的生长点[J]. 中国图书馆学报,2000,26(5):30-35.

[12] 李晨晖, 崔建明, 陈超泉. 大数据知识服务平台构建关键技术研究[J]. 情报资料工作, 2013(2):29-34.

[13] 王知津, 于晓燕. 数字图书馆网格的资源发现方法及其分析评价[J]. 图书馆杂志, 2010,29(6):16-20.

[14] 谢晓阳. 基于网格的知识服务平台[J]. 情报理论与实践, 2006,29(3):374-375,348.

[15] 高东平, 方安, 李杨, 等. 知识服务平台的设计与应用——以重大传染病信息知识服务平台为例[J]. 情报理论与实践, 2011,34(7):111-115.

[16] 姜永常. CNKI数字图书馆知识服务研究[J]. 情报学报, 2004,23(3):265-274.

[17] 马捷, 吴琼, 亓莉莉. 公共服务管理平台知识服务内容解析[J]. 图书情报工作, 2010,54(6):40-43,101.

[18] 吴赟, 孙梦如. 中国出版业发展知识服务的路径思考——从爱思唯尔ClinicalKey超级医学平台谈起[J]. 出版广角, 2017(13):13-16.

[19] 于彤, 张竹绿, 贾李蓉. 面向循证医学的知识服务平台概述[J]. 中国中医药图书情报杂志, 2014,38(4):55-57.

[20] BELKIN N J, MARCHETTI P G, ALBRECHT M, et al. User interfaces for information systems [J]. Journal of information science, 1991,17(6):327-344.

[21] 陈代春. 高校图书馆知识服务平台构建探析[J]. 情报科学, 2008,26(6):841-844.

[22] 何绍华, 王培林. 知识管理环境下的我国图书馆知识服务模式研究[J]. 情报理论与实践, 2007(5):619-621,637.

[23] 李芳慧, 王玲. "互联网+"时代下的知识服务平台运营模式探究[J]. 图书馆学研究, 2018(12):63-67.

[24] 常红, 田野, 沙淑欣. 图书馆信息推送服务用户接受研究[J]. 图书馆论坛,2014,34(11):106-111.

[25] 刘明珠, 杨建林. 微博搜索、网页搜索对用户信息需求满足能力的对比分析——以新浪微博搜索、百度搜索为例[J]. 图书与情报, 2016(5):126-136.

[26] 王若佳, 李月琳. 基于用户体验的健康类搜索引擎可用性评估[J]. 图书情报工作, 2016,60(7):92-102.

[27] 冯缨, 张瑞云. 基于用户体验的微博信息质量评估研究[J]. 图书馆学研究, 2014(9):62-67,101.

[28] 王璟璇, 杨道玲. 基于用户体验的政府网站绩效评估:探索与实践[J]. 电子政务, 2014(5):35-41.

[29] 陈丛聪. 面向用户体验的移动B2C界面可用性评估研究[D]. 重庆:西南大学, 2014.

[30] 齐雪. 数字图书馆交互功能评估研究[D]. 天津: 南开大学, 2014.

[31] BORLUND P, INGWERSEN P. The development of a method for the evaluation of interactive information retrieval systems [J]. Journal of documentation, 1997, 53(3):225-250.

[32] 李月琳, 肖雪, 胡蝶. 信息检索实验中的任务设计——真实与模拟仿真工作任务的比较研究[J]. 图书情报工作, 2014,58(16):5-12.

[33] LI Y, HU D. Interactive retrieval using simulated versus real work task situations: differences in sub-facets of tasks and interaction performance [J/OL]. Proceedings of the American Society for Information Science and Technology, 2013, 50(1):1-10. [2018-12-06]. http://www.asis.org/asist2013/proceedings/openpage.html.

作者贡献说明:

李月琳:负责选题、论文撰写及修改;

韩宏亮:负责文献收集、整理和格式调整。

李月琳，韩宏亮. 从信息检索系统评估到知识服务平台评估[J]. 图书情报工作,2019,63(1):52-59.

From the Evaluation for Information Retrieval Systems to the Evaluation for Knowledge Platform

Li Yuelin　Han Hongliang

The Department of Information Resources Management, Business School, Nankai University, Tianjin 300071

Abstract: [**Purpose/significance**] With the development of knowledge platforms, it has been an important issue to improve the function and design of knowledge platform through systems evaluation. The article is aimed at exploring the evaluation elements and criteria for the evaluation of knowledge platform. [**Method/process**] The article is start up with the evaluation of information retrieval. It reviews related studies on the evaluation of information retrieval systems and digital libraries, especially addressing the theoretical models, elements, and criteria of digital library evaluation. [**Result/conclusion**] Based on the review and discussion, the article identifies and addresses the elements and criteria for the evaluation of knowledge platforms.

Keywords: knowledge service　evaluation for knowledge platform　conceptual framework

《图书情报工作》2018 年优秀审稿专家

2018 年,有近 300 位审稿专家参加了《图书情报工作》稿件的同行评议工作,共评审稿件 1 400 余篇次,审阅 6 篇及以上的专家有 100 余位。高效、高质量的评审为《图书情报工作》遴选高质量稿件提供了保障。综合考虑今年以来的审稿数量、质量和时效,评选出 50 位 2018 年优秀审稿专家(名单如下)。《图书情报工作》为优秀审稿专家颁发证书并免费赠送一年期刊的电子版。感谢所有审稿专家对《图书情报工作》的大力支持!

(以下优秀审稿专家按姓名拼音排序):

姓名	工作单位
白如江	山东理工大学科技信息研究所
曹锦丹	吉林大学公共卫生学院
崔宇红	北京理工大学图书馆
邓胜利	武汉大学信息管理学院
邓小昭	西南大学计算机与信息科学学院
范爱红	清华大学图书馆
方向明	上海大学图书馆
冯　佳	上海社会科学院文学研究所
甘春梅	中山大学
高　凡	西南交通大学图书馆
韩　毅	西南大学计算机与信息科学学院
胡昌平	武汉大学信息资源研究中心
黄　崑	北京师范大学政府管理学院
黄国彬	北京师范大学政府管理学院
黄令贺	河北大学管理学院
姜春林	大连理工大学人文与社会科学学部科学学与科技管理研究所
李　刚	南京大学信息管理学院
李　明	南京大学信息管理学院
李　武	上海交通大学媒体与设计学院
李书宁	北京师范大学图书馆
李卓卓	苏州大学
刘　冰	天津师范大学管理学院
刘　勘	中南财经政法大学信息与安全工程学院
刘春丽	中国医科大学图书馆
刘晓娟	北京师范大学政府管理学院
刘雪立	新乡医学院期刊社/河南省科技期刊研究中心
刘兹恒	北京大学信息管理系
马　捷	吉林大学管理学院
马学良	国家图书馆
茆意宏	南京农业大学信息科学技术学院
牟冬梅	吉林大学公共卫生学院
裴　雷	南京大学信息管理学院
秦　鸿	电子科技大学图书馆
任树怀	上海外国语大学图书馆
邵　波	南京大学图书馆
滕广青	东北师范大学信息科学与技术学院
王立学	中国科学技术信息研究所
王晰巍	吉林大学管理学院
王延飞	北京大学信息管理系
吴　红	山东理工大学科技信息研究所
吴振新	中国科学院文献情报中心
向桂林	中国科学院生物物理研究所
闫　慧	中国人民大学信息资源管理学院
杨新涯	重庆大学图书馆
张广钦	北京大学信息管理系
张鹏翼	北京大学信息管理系
张卫东	吉林大学管理学院
赵　飞	北京大学图书馆
赵宇翔	南京理工大学
祝忠明	中国科学院兰州文献情报中心/中国科学院资源环境科学信息中心